Steven L. Beshear
Governor

Terry Holliday, Ph.D.
Commissioner of Education

EDUCATION AND WORKFORCE DEVELOPMENT CABINET
DEPARTMENT OF EDUCATION

Capital Plaza Tower • 500 Mero Street • Frankfort Kentucky 40601
Phone: (502) 564-4770 • www.education.ky.gov

September 13, 2012

President David Williams, LRC Co-Chair
Speaker Greg Stumbo, LRC Co-Chair
Legislative Research Commission
Capitol Building, Room 300
Frankfort, KY 40601

Dear Senator Williams and Representative Stumbo:

The Kentucky Department of Education previously transmitted to the Legislative Research Commission the 2010-12 biennial plan for validation studies as required by 158.6453(17). Please find enclosed a summary of the findings for the (1) 2010-12 biennium (Spring 2010 KCCT assessment and 2011 KCCT assessment); (2) a 2009 report completed after the previously submitted 2008-10 biennium reports; and, (3) reports for preliminary work on the Kentucky Performance Rating for Educational Progress (K-PREP) tests, the new public school college and career ready assessment system that the Kentucky General Assembly enacted into legislation in 2009 per Senate Bill 1. The complete reports are available on the Kentucky Department of Education website at www.education.ky.gov.

Please let us know if you have questions about the enclosure.

Sincerely,

Terry Holliday, Ph.D.

C: Ben Boggs
Ken Draut
Kentucky Board of Education members
Anita Muckelroy
Robert Sherman
Marcia Seiler
Ken Warlick

Enclosure

Interim Accountability Testing System: Validation and Research Agenda

Validation Studies and Evidence for Reliability

Kentucky Department of Education
Terry Holliday, Ph.D., Commissioner

September 2012

Kentucky
UNBRIDLED SPIRIT

Validation Studies and Evidence for Reliability

According to KRS 158.6453 (17), "the Department of Education shall gather information to establish the validity of the assessment and accountability program. It shall develop a biennial plan for validation studies that shall include, but not be limited to, the consistency of student results across multiple measures, the congruence of school scores with documented improvements to instructional practice and the school learning environment, and the potential for all scores to yield fair, consistent, and accurate student performance level and school accountability decisions. Validation activities shall take place in a timely manner and shall include a review of the accuracy of scores assigned to students and schools, as well as of the testing materials. The plan shall be submitted to the Commission by July 1 of the first year of each biennium. A summary of the findings shall be submitted to the Legislative Research Commission by September 1 of the second year of the biennium."

Scope of Work for Interim Accountability Testing System Research

The complete Scope of Work for the Office of Assessment and Accountability, Kentucky Department of Education (KDE), was reviewed by the Kentucky Board of Education (KBE), and was submitted to the Legislative Research Commission in July 2010. The plan was presented to the National Technical Advisory Panel on Assessment and Accountability (NTAPAA) for committee recommendations and review.

The 2010-12 biennium is a transition period moving from the Kentucky Core Content Test (KCCT) to the Kentucky Performance Rating for Educational Progress (K-PREP), the new College and Career Ready assessment system. During the Interim Accountability window, KCCT continued to appear on many documents.

Research activities described in the Scope of Work are undertaken to provide evidence as required by the statute described above. These research activities cover a broad range and include the following topics:

1. Annual Third-Party Checking of KCCT Scaling and Equating
   - *1.1* Third-Party Checking of 2011 Calibration and Scaling for the Kentucky Core Content Test (Document FR-11-65)
   - *1.2* Third-Party Checking of 2010 Calibration and Scaling for the Kentucky Core Content Test (Document FR-10-62)
   - *1.3* Third-Party Checking of 2009 Calibration and Scaling for the Kentucky Core Content Test (Document FR-09-79)

2. Preliminary Work on the Kentucky Performance Rating for Educational Progress (K-PREP) Tests
   - 2.1 An Exploration of Alternate Methods for Scoring and Estimating Item Parameters for the Kentucky Writing Assessments (Document FR-10-14)
   - 2.2 Alignment of Kentucky Core Content Test (KCCT) Items to the Common Core State Standards (Grades 6, 7, and 8) (Document FR-10-36)

2.3 Alignment of Kentucky Core Content Test (KCCT) Items in Reading and Mathematics to the Common Core State Standards: Grades 3 through 5 (Document FR-10-44)
2.4 Estimation of Students' QualityCore® End-of-Course Exam Grades (Document FR-11-52)

Detailed Description of Each Study

A more detailed description of each study is presented in the report. Each study is described in terms of:

A. Study Source (Who conducted the study?)
B. Purpose (Why do the research?)
C. Audience (Who will use the results of the research and how will they use it?)
D. Methodology (How will the research be conducted?)
E. Findings and Recommendations

1. Annual Third-Party Checking of KCCT Scaling and Equating

Study Source

The study is happening under contract with the Human Resources Research Organization (HumRRO).

Purpose

According to the request for proposals from which this study originated, "The Kentucky Department of Education is charged with "maintaining a vigorous ongoing program of research and documentation of the effects of the assessment and accountability system on Kentucky schools." This consequential validity study will examine the impact of the assessment system in two important dimensions. Tests of student achievement should be instructionally sensitive. Improving instruction should improve students' performance. Both the Interim Accountability Testing System and No Child Left Behind Act of 2001 (NCLB) accountability systems are predicated on the assumption that teachers and schools can improve student test scores by improving instruction. The first dimension of inquiry relates improved instructional practice to high or improving test scores.

In addition to the impact of instruction on test scores, there also exists the potential for the test (and the accountability system of which it is a part) to effect changes in instructional practice. The tested content has great potential to impact the curriculum taught within schools. The test item format might impact specific teaching practices. The grades at which particular subjects are tested can impact school schedules and staff assignments. Some responses by districts, schools, and teachers because of the testing system are expected, and in some instances, encouraged. Other responses might not be anticipated or considered positive in terms of the overall educational experience of Kentucky's students. For example, a school might completely eliminate science from the middle school curriculum after Grade 7 in an effort to concentrate on the subjects tested in Grade 8. It is important to document the existence and extent of positive anticipated consequences to the

accountability system. It is also important to document unanticipated consequences so that we might better guard against negative responses.

Audience

This research will be used to inform policy stakeholders (KBE, KDE, NTAPAA, Office of Education Accountability (OEA)) of the link between Kentucky Common Core Academic Standards, K-PREP, instructional practice, and the learning environment. Since NTAPAA makes policy recommendations to the KBE, the report will be directed to NTAPAA, KBE, and KDE.

Methodology

Every year since 1998[1], the KCCT has been scaled and equated, and then raw-score-to-scale-score tables have been produced to be applied to students' test results. Item parameters for the current year's test are estimated using Item Response Theory (IRT). These item parameters are linearly transformed (scaled) and equated (linked) with previous years' scales. The results of scaling and equating are then used to construct raw-score-to-scale-score tables for every KCCT test form. Cut points are also identified so that students' scores can be translated to performance categories: Novice, Apprentice, Proficient, and Distinguished (NAPD).

As a quality control step, the testing contractor's psychometric staff and personnel at HumRRO conduct scaling and equating analyses simultaneously and independently. Researchers at both KDE and HumRRO compare results at several steps throughout the process. If a result between the two companies is not identical, then procedures are reviewed until the issue is resolved and both staffs get the same outcome. In this way, the item parameter estimation analyses, scaling, equating, raw-score-to-scale-score transformations, and cut point identifications are checked and verified by two autonomous agencies. HumRRO has served as the third-party checker or as the primary psychometric contractor for the KCCT since the 1998–99 academic year. The results presented in this report are comparable to prior third-party investigations of KCCT psychometric processing.

In addition to performing a parallel analysis with Measured Progress, HumRRO also conducts an in-house parallel analysis to accomplish scaling, equating and the production of raw-score-to-scale-score tables.

1.1 Third-Party Checking of 2011 Calibration and Scaling for the Kentucky Core Content Test

Findings and Recommendations

Two independent psychometric teams, one from Measured Progress and one from Human Resources Research Organization (HumRRO), independently estimated item parameters, scaled, and equated the 2011 Kentucky Core Content Test (KCCT). They then verified that the 2009 scoring tables could be applied to the 2011 KCCT administration for

[1] The test in use before 1998 was the Kentucky Instructional Results Information System (KIRIS) test.

mathematics, reading, science and social studies. New scoring tables were produced for writing. Procedures for scaling and equating were agreed upon by Measured Progress and HumRRO before processing began. The KCCT for reading, mathematics, science and social studies used the same items as the 2009 administration; as a result, the psychometric process for 2011 was to verify that the 2009 scoring tables could be applied to the 2011 administration. Several procedures were conducted independently by Measured Progress and HumRRO to determine the appropriateness of applying the 2009 scoring tables to the 2011 student scores. Decisions regarding the handling of discrepancies that arose during the process of parameter estimation, scaling, equating, and the production of scoring tables were discussed between Measured Progress and HumRRO, and in all cases both groups reached consensus. Ultimately, HumRRO's recommendations to apply the 2009 scoring tables to the 2011 KCCT administration conceded with Measured Progress's recommendations. HumRRO matched Measured Progresses results for writing. Thus, HumRRO is assured that Measured Progress did not commit processing errors for writing and that they appropriately applied the 2009 scoring tables for mathematics, reading, science and social studies to the 2011 test administration.

1.2 Third-Party Checking of 2010 Calibration and Scaling for the Kentucky Core Content Test

Findings and Recommendation

Two independent psychometric teams, one from Measured Progress and one from HumRRO, independently estimated item parameters, scaled, and equated the 2010 Kentucky Core Content Test (KCCT). They then produced the raw-score-to-scale-score tables to be applied to students' test results. The 2010 KCCT was equated to the prior KCCT scale for all grade/subjects. Procedures for equating were agreed upon before processing began by Measured Progress and HumRRO with input and advice from KDE. For Science and Social Studies the 2010 KCCT was equated using items administered in 2008 or 2009. The KCCT for reading Grades 6 and 7 and mathematics Grades 3, 4, 6, and 7 were equated using only parameters from the 2008 administration. Nearly all the items for those grade/subjects were also administered in 2008. For reading Grades 3, 4, 5, 8 and 10 and mathematics Grades 5, 8 and 11, the 2008 raw-score-to-scale-score tables were applied to the current raw scores. The items on those assessments were exactly the same as in 2008, so no equating was necessary. Decisions regarding the handling of discrepancies that arose during the process of parameter estimation, scaling, equating, and the production of raw-score-to-scale-score tables were discussed between Measured Progress and HumRRO, and in all cases both groups reached consensus. Ultimately, the 2010 results calculated by HumRRO were identical to those calculated by Measured Progress. Given that HumRRO's results were identical with those of Measured Progress, HumRRO is assured that Measured Progress did not commit processing errors.

1.3 Third-Party Checking of 2009 Calibration and Scaling for the Kentucky Core Content Test

Findings and Recommendations

Measured Progress and HumRRO independently scaled and equated the Kentucky Core Content Test (KCCT), and then produced the raw-score-to-scale-score tables to be applied to students' test results. The 2009 KCCT was equated to the 2007 KCCT results for high school reading, mathematics, and science. All other grades and subjects were equated to the 2008 KCCT results. Decisions regarding the handling of discrepancies that arose during the process of scaling, equating, and the production of raw-score-to-scale-score tables were discussed between Measured Progress and HumRRO, and in all cases both groups reached consensus. As a result of those discussions, a new procedures document for scaling and equating was developed by Measured Progress and approved with revisions by HumRRO and KDE. This newly developed procedures document will guide scaling and equating for 2010. Ultimately, the 2009 results calculated by HumRRO were identical to those calculated by Measured Progress. Given that HumRRO's results were identical with those of Measured Progress, HumRRO is assured that Measured Progress did not commit processing errors.

2. Preliminary Work on the Kentucky Performance Rating for Educational Progress (K-PREP) Tests

2.1 An Exploration of Alternate Methods for Scoring and Estimating Item Parameters for the Kentucky Writing Assessments.

Study Source

The study is happening under contract with the Human Resources Research Organization (HumRRO).

Purpose

In 2007, Kentucky shifted the scoring of their on-demand writing assessment from a holistic to an analytic method. Where students previously received only an overall NAPD classification, they now receive a scaled score and analytic/diagnostic information designed to indicate their strengths and weaknesses on four components of writing. Concerns regarding the scaling and equating of the writing assessment, as well as concerns related to the scoring patterns for the writing prompt, led to the investigations presented in this report.

Concerns about the writing prompt have led Kentucky to equate the on-demand writing assessment using only multiple-choice items. An investigation of the impact of that decision was conducted by equating using both multiple-choice items and the writing prompt and making direct comparisons regarding the classification of students. The inclusion of the writing prompt did change the classification outcome for some students. The change was small; typically 1% or fewer students changed classification for Grade 8, and the largest change in classification for Grade 5 was about 3%. These small differences indicate that the decision to omit the prompts from equating does not substantially alter the outcome of the assessment.

Audience

This information will facilitate judgment of test forms comparability by KDE, KBE, the NTAPAA, and other stakeholders. The test-construction contractor will gain feedback for use in subsequent test construction.

Methodology

This study compared multiple-choice equating with equating that included the writing prompt scored for content only. Because the writing prompt is also scored for structure and conventions, it is possible that the equating might have turned out somewhat differently had we chosen one of the other scored components. By mutual decision between Measured Progress, KDE and HumRRO, because the IRT parameter estimation software would not generate plausible parameters for all three highly correlated components, the item parameters for the writing prompt have been generated from the content component only. These were therefore the parameters available for use in equating and for this study. Other evidence from this study indicates that the parameters may have exhibited greater differences had one of the other components been chosen.

The next investigation examined the differences that might have occurred had Kentucky chosen a different component to score on the writing prompt. To examine the potential for differences, each component was used separately to generate parameters. Then, three separate raw-score-to-scale-score tables were produced using the multiple-choice items plus each prompt score component. This had the overall effect of reducing the total raw score from 48 on the operational test to 32 for each prompt component test. The reduction in numbers of score points was expected to cause some differences in the classifications of students.

There were some differences in classification depending on the component score that was used. Classifications were very close when results were generated using only the content component or the structure component. However, classification differed the most from operational scores when the conventions component was used. Overall, results suggest that adding two components to students' writing scores without generating parameters to describe those components may create noise in the measurement. Given the high correlations, it is difficult to support the idea that the three components are measuring substantially different content.

Similar writing assessments have been more successful at differentiating among the specified subcontent areas than Kentucky's on-demand writing assessment. Roid (1994) reported that Oregon was able to generate six separate dimensions of writing, including ideas, organization, voice, word choice, sentence fluency, and conventions. The six dimensions were positively correlated between .49 and .78. Despite these relatively weaker correlations, it was also reported that some of the dimensions could have been combined with little loss of information. Lane's (2006) review of the literature suggests that analytic rubrics typically have the potential to produce distinct information for only a small number (i.e., 2 or 3) of domains.

Findings and Recommendations

For equating purposes, Kentucky has chosen to use only multiple-choice items. This report shows that only small differences in categorizations would occur if prompt scores based on content had been included. Should Kentucky desire to include prompt scores for operational equating, it is recommended that revisions to the scoring of the prompts be made first. Using the prompt scores for equating may bolster the accuracy of the equating solution by adding a prompt component to the process component measured by the multiple-choice items. However, the manner in which the prompt score is included in the equating solution should be deliberate, should not include any replication of item parameters across components, and should be based on the best possible estimate of student achievement on a well-defined identifiable writing construct.

Since Kentucky's on-demand writing assessment produces such highly correlated component scores for the writing prompt, there is not sufficient support to justify continuing to produce student subscores based on these data. The use of a single set of item parameters to represent all three components would not have been justifiable had the content been substantially different. The subscores, at best, provide redundant information and, at worst, add noise to the measurement, potentially promoting poor instructional decisions.

Therefore, HumRRO recommends that students be assigned a single score and a single proficiency category for writing barring significant changes to the assessment structure, the scoring methods, and/or the scoring rubrics. If the writing construct is to be divided into multiple components, it is recommended that the decision to do so be informed by indications of convergent/discriminant validity (correlational evidence that different constructs are being measured) and factor analysis (modeling equations that indicate whether multiple factors are being measured by the writing assessment and how those factors are structured). If Kentucky chooses to continue to produce subscores for writing, these analyses will ensure that the subscores represent distinguishable knowledge and skills that can be translated to describe students' strengths and weaknesses among the identified subscores.

Until Kentucky can gather evidence for assessing multiple writing components, it is also recommended that the practice of assigning the same item parameters to all three writing components scored on the prompt be discontinued. Sufficient differences were found when generating distinct parameters for each component and applying them to student scores, calling this practice into question. It is suspected that this difference is because of a scale issue rather than because the students were rank ordered differently by component. Applying the same parameters across components may be adding measurement error to the overall writing scores. Kentucky might consider providing subscores based on the multiple-choice items (a process score) and an amalgam of the currently defined writing prompt components.

2.2 Alignment of Kentucky of Content Test (KCCT) Items to the Common Core State Standards (Grades 6, 7 and 8)

Study Source
The study is happening under contract with the Human Resources Research Organization (HumRRO).

Purpose
HumRRO conducted a test item alignment review of the KCCT items in Grades 6, 7, and 8 to the Common Core State Standards in mathematics and reading for KDE. The standards were adopted by KBE and are now referred to as the Kentucky Core Academic Standards. The purpose of this study was to evaluate where the current KCCT items fall relative to the Common Core Standards. The alignment review examined the degree of match to content categories, as well as depth of knowledge (DOK) expected in the Common Core Standards.

Audience
Results can be used to inform item development as Kentucky attempts to align the assessment system with Common Core. Pearson, the test-construction contractor, will gain feedback for use in subsequent test construction.

Methodology
HumRRO staff experienced with alignment research performed the item review and analyses. Staff reviewed the content area (reading or mathematics) in which they hold the most expertise. The review consisted of an evaluation of KCCT items in reading and mathematics for each of Grades 6, 7 and 8 relative to the new Common Core Standards in the respective content areas.

HumRRO evaluated alignment of KCCT reading and mathematics items to the Common Core State Standards by performing several tasks common to alignment methodologies. For the standards documents, reviewers collaboratively examined individual standards to determine the DOK expected for students to demonstrate content proficiency. For assessment items, reviewers rated items independently on two dimensions: (a) standard match to identify primary content targeted by item, and (b) DOK to determine the extent of processing needed to respond successfully to items.

HumRRO compared test items to the Common Core State Standards released June 2, 2010 in reading and in mathematics (http://www.corestandards.org/) to the current KCCT item pool for reading and mathematics in Grades 6, 7, and 8.

Findings and Recommendations
Regarding content match, overall the results indicate that the KCCT items in both reading and mathematics assess a variety of the Common Core Standards. Items do tend to cluster around some content more than others; thus, the range of content assessed relative to the Common Core Standards is somewhat narrow. In addition, while no Common Core strands or substrands are omitted entirely, some specific standards could not be linked to

any KCCT items; conversely, some KCCT items did not match any Common Core Standards.

Although content clustering and omission are concerns if Kentucky wishes to make use of the current item pools while connecting to the Common Core Standards, no *major* gaps exist. Furthermore, some of the discrepancies found between the item pools and content standards simply stem from the organization selected (i.e., where content expectations are placed) for the Common Core Standards document, which differs some from the organization of the Kentucky Core Content Academic Standards and that of many other states. For example, vocabulary knowledge and literary devices can be found in the Common Core English Language Arts standards; however, they exist under language and writing standards instead of reading.

For both reading and mathematics, a sizeable number of items were matched to Grade 6 Common Core Standards, even for Grade 7 and Grade 8 KCCT items. While this outcome does not necessarily indicate that Grades 7 and 8 curricula focus more on Common Core Standards than Grade 6, Kentucky should be aware of this feature of the Common Core State Standards.

In addition to evaluating content categories assessed, HumRRO examined the depth of processing in KCCT items relative to the level of processing expected in the Common Core Standards. The results for mathematics suggest that the consistency between the KCCT and Common Core Standards is good because reviewers determined that many items assessed students at the same cognitive level as the corresponding standards. Mathematics items matched to content expectations within several substrands did fall below standard, which may prompt Kentucky to review and potentially increase DOK for these items. For example, reviewers determined that the majority of items matched to the Statistics and Probability strand assessed students below the standard. However, some of the items were matched to standards above the grade level of the item, which may not be as critical as for on-grade items matched to on-grade standards. In addition, almost all discrepancies identified count as adjacent mismatches (i.e., item rated as DOK 1 and standard rated as DOK 2). In comparison, approximately half of KCCT reading items at each grade were rated at a different level than expected in the Common Core Standards.

One consideration for Kentucky in general regarding item-level depth-of-knowledge is to consider increasing the number of items assessing higher-order thinking (DOK levels 3 and 4) on the assessment. Working towards items assessing students at a higher cognitive level has been a national trend in the last several years beyond the Common Core State Standards effort.

2.3 Alignment of Kentucky of Content Test (KCCT) Items in Reading and Mathematics to the Common Core State Standards: Grades 3 through 5

Study Source
The study is happening under contract with the Human Resources Research Organization (HumRRO).

Purpose
HumRRO conducted a test item alignment review of the KCCT items in Grades 3, 4 and 5 to the Common Core State Standards in mathematics and reading for KDE. The purpose of this study was to evaluate where the current KCCT items fall relative to the Common Core Standards. The alignment review examined the degree of match to content categories, as well as depth of knowledge, expected in the Common Core Standards. Results can be used to inform item development as Kentucky attempts to align the assessment system with the Common Core.

Audience
This information can be used to inform item development as Kentucky attempts to align the assessment system with Common Core. The test-construction contractor will gain feedback for use in subsequent test construction.

Methodology
HumRRO staff experienced with alignment research performed the item review and analyses. Staff reviewed the content area (reading or mathematics) in which they hold the most expertise. The review consisted of an evaluation of KCCT items in reading and mathematics for each of Grades 3, 4 and 5 relative to the new Common Core Standards in the respective content areas.

HumRRO evaluated alignment of KCCT reading and mathematics items to the Common Core State Standards by performing several tasks common to alignment methodologies. For the standards documents, reviewers collaboratively examined individual standards to determine the depth of knowledge (DOK) expected for students to demonstrate content proficiency. For assessment items, reviewers rated items independently on two dimensions: (a) standard match to identify primary content targeted by item, and (b) DOK to determine the extent of processing needed to respond successfully to items.

Findings and Recommendations
Regarding content match, overall the results indicate that the KCCT items in both mathematics and reading assess a variety of the Common Core Standards. Items tend to cluster around some content more than others (i.e., Key Ideas and Details in Reading; Operations and Algebraic Thinking in Mathematics). While this outcome makes the range of content assessed relative to the Common Core Standards somewhat narrow, the emphasis on this content in particular among KCCT does seem to correspond with intended emphasis in the Grades 3, 4 and 5 Common Core Standards. KDE may wish to review specific Common Core Standards that could not be linked to any KCCT items, as well as those KCCT items not matched to any Common Core Standards.

Although content clustering and omission are concerns if Kentucky wishes to make use of the current item pools while connecting to the Common Core Standards, no *major* gaps

exist. Furthermore, some of the discrepancies found between the item pools and content standards simply stem from the organization selected (i.e., where content expectations are placed) for the Common Core Standards document, which differs some from the organization of the Kentucky Core Content Standards and that of many other states. For example, vocabulary knowledge and literary devices can be found in the Common Core English Language Arts Standards; however, they exist under language and writing standards instead of reading.

For both reading and mathematics, a number of items corresponded with Common Core Standards above and below the grade level in which they are assessed. In these cases, KDE may consider moving items to better match the Common Core organization.

In addition to evaluating content categories assessed, HumRRO examined the level of processing required to respond to KCCT items relative to the level of processing expected in the Common Core Standards. The results for mathematics suggest that the consistency between the KCCT and Common Core Standards is quite good overall because reviewers determined that many items assessed students at the same cognitive level as the corresponding standards. Mathematics items matched to content expectations within several strands did fall below and also above the corresponding standard, which may prompt Kentucky to review these items for greater consistency with the Common Core if necessary. Many of these discrepancies identified count as adjacent mismatches (i.e., item rated as DOK 1 and standard rated as DOK 2). In comparison, a substantial number of Grade 3 KCCT reading items, as well as almost all of Grade 4 items, were rated at a lower DOK level than expected in the Common Core Standards.

One consideration for Kentucky in general regarding item-level depth-of-knowledge is to consider increasing the number of items assessing higher-order thinking (DOK levels 3 and 4) on the assessment. Working towards items assessing students at a higher cognitive level has been a national trend in the last several years beyond the Common Core State Standards effort.

2.4 Estimation of Students' QualityCore® End-of-Course Grades

Study Source

The study is happening under contract with the Human Resources Research Organization (HumRRO).

Purpose

In the 2011-12 academic year Kentucky high school students will take end-of-course (EOC) examinations (ACT's QualityCore®) Exams) in Algebra II, Biology, English II (Sophomore English), and U.S. History. These courses are designed to ascertain if a student has mastered the content sufficiently to move to the next course in the series. Kentucky will also encourage teachers to use the exams as a component of student course grades. This study was designed to provide guidance on how students' grades could be assigned based on EOC exam scores.

Audience

This information can be used to inform policy stakeholders (the KBE, Kentucky Department of Education (KDE), National Technical Advisory Panel on Assessment and Accountability (NTAPAA)) to guide teachers on how students' grades can be assigned based on End-of-Course (EOC) test scores.

Methodology

This study begins by establishing a single "cut score" on the exams. This cut score should coincide with some meaningful categorization or description of students' achievement in the course. KDE is fortunate in this regard because the Kentucky Council on Postsecondary Education (CPE) has adopted the ACT assessment for determining college readiness and set cut scores on the ACT to indicate that students are ready to take credit bearing college courses. HumRRO began this study with the ACT cut scores, and linked them to scores on the QualityCore® exams. For this study, those cut scores are as follows:

Table 1. ACT College Readiness Cut Scores Set by CPE

Content Area	ACT Score
Reading	20
English	18
Mathematics	19
Science*	19

*CPE did not indicate a science cut score. The average of the other subjects (19) was used for this study.

The report continues to detail the steps using ACT and PLAN results to develop cut points for EOC grades,

Findings and Recommendations

Obviously, this represents only one of the multitude of solutions that could be applied to setting cut scores for grading the EOC exams. KDE might assign different percentages or effect sizes to separate one grade from another. KDE might choose a different part of the ACT benchmark linked range as the starting point.

This solution has the following advantages:

- EOC scores are linked to an important college readiness indicator.
- The procedure was applied in the same manner across all subjects.
- The solutions are relatively easy to explain to the field.
- Kentucky educators provided input and guidance during the process.

Final Recommended Scoring Table

Subject	A	B	C	D	F	CPE-linked College Readiness Benchmark
US History	153-above	148-152	144-147	136-143	135-below	148
English II	158-above	150-157	144-149	136-143	135-below	150
Algebra II	149-above	144-148	141-143	136-140	135-below	144
Biology	153-above	148-152	145-147	136-144	135-below	148